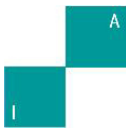
Inteligencia Artificial 20(59) (2017), 70-81
doi: 10.4114/intartif.vol20iss59pp70-81



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Rational versus Intuitive Outcomes of Reasoning with Preferences: Argumentation Perspective

Kristijonas Čyras
Imperial College London
k.cyras13@imperial.ac.uk

Abstract Reasoning with preference information is a common human activity. As modelling human reasoning is one of the main objectives of AI, reasoning with preferences is an important topic in various fields of AI, such as Knowledge Representation and Reasoning (KR). Argumentation is one particular branch of KR that concerns, among other tasks, modelling common-sense reasoning with preferences. A key issue there, is the lack of consensus on how to deal with preferences. Witnessing this is a multitude of proposals on how to formalise reasoning with preferences in argumentative terms. As a commonality, however, formalisms of argumentation with preferences tend to fulfil various criteria of 'rational' reasoning, notwithstanding the fact that human reasoning is often not 'rational', yet seemingly 'intuitive'. In this paper, we study how several formalisms of argumentation with preferences model human intuition behind a particular common-sense reasoning problem. More specifically, we present a common-sense scenario of reasoning with rules and preferences, complemented with a survey of decisions made by human respondents that indicates an 'intuitive' solution, and analyse how this problem is tackled in argumentation. We conclude that most approaches to argumentation with preferences afford a 'rational' solution to the problem, and discuss one recent formalism that yields the 'intuitive' solution instead. We argue that our results call for advancements in the area of argumentation with preferences in particular, as well as for further studies of reasoning with preferences in AI at large.

Keywords: Knowledge Representation and Reasoning, Argumentation, Preferences

1 Introduction

Preferences are quotidian phenomena. They appear in everyday reasoning problems, within utterances such as "I prefer Brownie over Almond cake". Since AI is inherently concerned with human type-of reasoning, preferences are of natural interest to the AI community. Lately, dealing with preference information has increasingly become of great interest in areas of AI that involve reasoning under uncertainty (see e.g. [15, 23, 33, 42, 43] for discussions).

Argumentation (as overviewed in [43]) is by now an acknowledged branch of one of the main subfields of AI, namely Knowledge Representation and Reasoning (KR), that is particularly concerned with reasoning with incomplete, uncertain and conflicting information. It is widely used in areas such as multi-agent systems, common-sense reasoning, decision making (see e.g. [2, 18, 33, 36, 43]). Preferences in argumentation, as well as in AI on the whole, are used to, for example, qualify the uncertainty of, or discriminate over, information. A principal issue, however, is the lack of consensus on how preferences should be accounted for (see e.g. [33] for a relatively recent overview). Hence the variety of formalisms of argumentation with preferences: e.g. [1, 3, 5, 7, 11, 14, 21, 22, 26, 28, 30, 34, 38, 40, 41, 45, 46].

In this paper we propose a simple common-sense reasoning scenario concerning decision making in the presence of preference information. The scenario is accompanied by a social experiment, i.e. a survey, that indicates the intuitive decision that people make when faced with the problem in question. By investigating how current approaches to argumentation with preferences handle the scenario, we find that all of them, except the formalism recently proposed in [21, 22], yield different outcomes than human intuition dictates.

ISSN: 1988-3064(on-line)
©IBERAMIA and the authors

The scenario is described as follows.

Example 1 (Cakes)**.** There are three pieces of cakes on a table: a piece of Almond cake, a Brownie, and a piece of Cheesecake. You want to get as many cakes as possible, and the following are the rules of the game.

a) You can take cakes from the table in two 'rounds':
   1. In the first round you can take at most two cakes;
   2. In the second round you can take at most one cake.

b) If you take Almond cake and Cheesecake in the first round, Brownie will not be available in the second round. (Nothing is known about other possible combinations.)

Finally, very importantly, suppose that you *prefer* Brownie over Almond cake. (No other preferences.)
Which pair(s) of cakes would you choose in the first round?

This example falls into the family of reasoning (or, decision making) problems that involve reasoning with *rules* and *preferences*. In this case, there is essentially a single rule, namely that taking Almond cake and Cheesecake removes Brownie.[1] There is also only one preference, namely that Brownie is preferred over Almond cake. Given this information, what could be the reasoning outcome?

Obviously, {Almond cake, Cheesecake} is not a good option, because it prevents one from getting the Brownie. Meanwhile, among pairs {Almond cake, Brownie} and {Brownie, Cheesecake}, both result into the desired outcome. In addition, neither of the two is in conflict with the preference information: on the one hand, since there are no preferences involving Cheesecake, {Almond cake, Brownie} does not violate the preference of Brownie over Almond cake; on the other hand, {Brownie, Cheesecake} satisfies the preference too, as along Cheesecake, the more preferred item is chosen instead of the less preferred one.

So both choices {Almond cake, Brownie} and {Brownie, Cheesecake} seem to be equally good, or 'rational', in the sense that both lead to the desired outcome of obtaining all the cakes and both satisfy the preference relation. As a decision, one could thus randomly choose between the two pairs.

However, upon the inception of the problem, it seemed as if the {Brownie, Cheesecake} choice is somehow more intuitive, or preferred.[2] We have therefore, with the hypothesis in mind that {Brownie, Cheesecake} is the preferred choice, conducted an anonymous survey with precisely the formulation as in Example 1. There were four possible answers:

- {Almond cake, Brownie}
- {Brownie, Cheesecake}
- Indifferent between {Almond cake, Brownie}, and {Brownie, Cheesecake}.
  (I.e. randomly choose one of the two pairs.)
- Other

The three concrete answer choices were randomized for each respondent, while the 'Other' choice, which allowed for a specification, was always the last one.

In the survey, out of total 84 participants, 41 (i.e. 48.81 %) chose "Brownie and Cheesecake", and only 12 (11.90 %) chose "Almond cake and Brownie", while among the rest, 32 (38.10 %) participants said they were indifferent between the two pairs, and 1 (1.19 %) person suggested taking Almond cake and Cheesecake.[3] So, not only that {Brownie, Cheesecake} dominated more than four times {Almond cake, Brownie}; it also got 10 % more responses than the option of being indifferent between the two pairs. This suggests that Brownie and Cheesecake is the more 'intuitive' choice in this problem.

There may presumably be multiple explanations for the mismatch between the results of the survey and the 'rational' solution delineated above. We are, however, not going to speculate on this issue; there are numerous accounts, e.g. [35, 37] to name a few, on human (ir)rationality in, for instance, decision making, game theory, as well as argumentation itself. Rather, we will attempt to present formalizations of the problem in question in various approaches to argumentation with preferences. The analysis will reveal that current formalisms opt mostly for the normative 'rational' solution (i.e. either of the two pairs), rather than the observed 'intuitive' one (Brownie and Cheesecake). At the end of our exposition, we will also encounter a recently proposed formalism that does opt for the 'intuitive' solution, at least in the Cakes problem.

[1] Since all three cakes are wanted, condition a) simply indicates that the first round decision is pivotal and essentially revolves around only one rule expressed in condition b).

[2] This was initially confirmed by posing the problem to a handful of colleagues.

[3] The survey results can be found at https://www.surveymonkey.com/results/SM-GLNNBZ8Q/ . We have surveyed PhD students at the Department of Computing, Imperial College London. Invitations to take the survey were distributed by email via the Department's PhD students' mailing list. 79 responses were obtained in a single day. We do not claim any statistically significant findings; the survey was meant to support/undermine our belief in the way how preferences work in common-sense reasoning problems.

2 Cakes Example in Argumentation

In argumentation information is represented via *arguments* and *attacks* among them. For instance, abstract argumentation (AA) [25] frameworks are tuples $(Args, \rightsquigarrow)$ with a set $Args$ of arguments and a binary attack relation $\rightsquigarrow$ on $Args$: for $\mathsf{A}, \mathsf{B} \in Args$, we say that A attacks B just in case $\mathsf{A} \rightsquigarrow \mathsf{B}$. Whereas in AA the internal structure of arguments is unknown, structured argumentation (see e.g. [9] for a recent overview) allows for more granularity in defining arguments and attacks. Usually, a formal language for representing knowledge is assumed, and arguments are constructed as deductions from premises to conclusions, with attacks constructively defined.

For both abstract and structured argumentation, *semantics* is an essential procedure for designating sets $E \subseteq Args$ of arguments which can be deemed collectively acceptable. Following [25], we can summarize the notions required to define the semantics used in this paper thus.

Definition 1. Given an abstract argumentation framework $(Args, \rightsquigarrow)$, a set $E \subseteq Args$ of arguments is:

- **conflict-free** just in case for no arguments $\mathsf{A}, \mathsf{B} \in E$ it holds that $\mathsf{A} \rightsquigarrow \mathsf{B}$;
- **admissible** just in case E is conflict-free and for each argument $\mathsf{A} \in Args$ such that $\mathsf{A} \rightsquigarrow \mathsf{B}$ for some $\mathsf{B} \in E$, there is some $\mathsf{C} \in E$ with $\mathsf{C} \rightsquigarrow \mathsf{A}$;
- **preferred** just in case E is $\subseteq$-maximally admissible;
- **stable** just in case E is conflict-free and for each $\mathsf{A} \notin E$ there is $\mathsf{B} \in E$ with $\mathsf{B} \rightsquigarrow \mathsf{A}$.

When E is preferred, respectively stable, we also call E a preferred, respectively stable, *extension*.

For instance, if $Args = \{\mathsf{A}, \mathsf{B}\}$ and $\mathsf{B} \rightsquigarrow \mathsf{A}$ is the only attack, then $\{\mathsf{B}\}$ is a unique preferred/stable extension.

Many different argumentation semantics have been investigated (see e.g. [4, 43] for overviews). Broadly speaking, they represent different modes of reasoning with respect to certainty of information, i.e. more sceptical, or more credulous. Almost all of the semantics respect the cornerstone principles of conflict-freeness and admissibility, defined above. Preferred and stable semantics are the principal ones among credulous semantics [8]; we therefore focus on these two semantics in this paper, noting in advance, however, that distinction between the two will be hardly visible, as in almost all cases extensions under the two semantics coincide (but see Subsect. 2.5). Sceptical semantics, most notably the grounded semantics [25], will not be of interest because with our Cakes example it provides either empty solutions (Subsect.s 2.1, 2.2, 2.4), or coincides with the preferred and stable semantics (Subsect. 2.3). (An interested reader can consult the literature and verify this.)

A commonality of majority of argumentation formalisms that deal with preference information is to use preferences to modify the attack relation: attacks from less preferred arguments are discarded, i.e. if B is less preferred than A (in symbols, $\mathsf{B} \lhd \mathsf{A}$),[4] then $\mathsf{B} \rightsquigarrow \mathsf{A}$ fails; see e.g. [1, 7, 11, 14, 26, 28, 30, 34, 41]. Some formalisms (e.g. [45, 46]) employ preferences to select among extensions the 'preferable' ones, or even combine the two uses (e.g. [3]). In what follows we select several representatives of the broad family of approaches, and show that in modelling the Cakes problem, they do not yield the 'intuitive' answer to choose Brownie and Cheesecake. This, we believe, illustrates that the task of capturing human reasoning in argumentation calls for novel approaches to preference handling. This call has recently been answered in [21], whereby the authors proposed a new approach to structured argumentation with preferences, one that opts for the 'intuitive' solution in Cakes problem. We discuss different approaches in turn.

2.1 Deductive Argumentation

We begin with a well known formalism called Deductive Argumentation [10, 11] in which classical logic is commonly used as a basis. Deductive Argumentation can be seen as a representative of those formalisms that employ forms of propositional, first-order, conditional or temporal logics, e.g. [1, 10, 11, 28, 31].

Assume classical propositional logic (PL) with atoms $\mathsf{a}, \mathsf{b}, \mathsf{c}, \ldots$, connectives $\neg, \wedge, \rightarrow$ (including falsum $\bot$) and the classical consequence relation $\vdash$. A knowledge base is a set Δ of (PL) formulas. Given Δ, for $A \subseteq \Delta$ and a formula α, $\langle A, \alpha\rangle$ is an *argument* if $A \vdash \alpha$, $A \nvdash \bot$, and there is no $A' \subsetneq A$ with $A' \vdash \alpha$. We say $\langle A, \alpha\rangle$ *attacks* $\langle B, \beta\rangle$, written $\langle A, \alpha\rangle \rightsquigarrow \langle B, \beta\rangle$, if $\alpha \vdash \neg\mathsf{b}$ for some $\mathsf{b} \in B$. Given a preference relation, i.e. a strict partial order, $\lhd$ over arguments, we say that $\langle A, \alpha\rangle$ *defeats* $\langle B, \beta\rangle$, written $\langle A, \alpha\rangle \hookrightarrow \langle B, \beta\rangle$, if $\langle A, \alpha\rangle \rightsquigarrow \langle B, \beta\rangle$ and $\langle A, \alpha\rangle \not\lhd \langle B, \beta\rangle$.

Deductive Argumentation is completely modular with respect to preference relations $\lhd$ over arguments and is not concerned with how they are obtained. Thus, given a preference relation (a strict partial order) $<$ over, say, atoms, one has to lift it up to the argument level. This is known as preference aggregation. For instance, we can employ versions of the well known Elitist and Democratic ordering principles (see e.g. [40]), respectively called Disjoint Elitist [47] and Disjoint Democratic orders, denoted by $\lhd_{DE}$ and $\lhd_{DD}$, and defined as follows:

[4] The most common approach to represent preferences in argumentation is by using a binary relation over a set of objects, such as arguments; details will be provided in the following subsections.

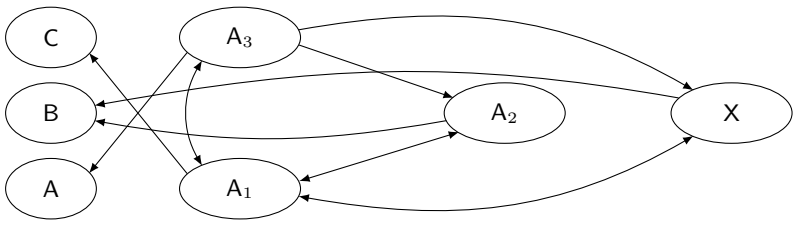
$\langle S, \sigma\rangle \lhd_{DE} \langle S', \sigma'\rangle$ iff $\exists \mathsf{s} \in S \setminus S'$ such that $\forall \mathsf{s}' \in S' \setminus S$ we have $\mathsf{s} < \mathsf{s}'$;
$\langle S, \sigma\rangle \lhd_{DD} \langle S', \sigma'\rangle$ iff $\forall \mathsf{s} \in S \setminus S'$ it holds that $\exists \mathsf{s}' \in S' \setminus S$ with $\mathsf{s} < \mathsf{s}'$.

Let us see how Deductive Argumentation equipped with these preference relations copes with our Cakes example.

Example 2 (Cakes in Deductive Argumentation)**.** Suppose that atoms $\mathsf{a}, \mathsf{b}, \mathsf{c}$ stand for the different pieces of cake, namely Almond cake, Brownie, Cheesecake, respectively. The rule that taking Almond cake and Cheesecake in the first round kicks out Brownie in the second one, can be represented as $\mathsf{a} \wedge \mathsf{c} \to \neg\mathsf{b}$. Finally, the preference of Brownie over Almond cake is specified as $\mathsf{a} < \mathsf{b}$. We thus have the knowledge base $\Delta = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{a} \wedge \mathsf{c} \to \neg\mathsf{b}\}$ with preference $\mathsf{a} < \mathsf{b}$. The following arguments can then be constructed:[5]

$\mathsf{A} = \langle\{\mathsf{a}\}, \mathsf{a}\rangle$, $\mathsf{A}_1 = \langle\{\mathsf{a}, \mathsf{b}\}, \mathsf{a} \wedge \mathsf{b}\rangle$,
$\mathsf{B} = \langle\{\mathsf{b}\}, \mathsf{b}\rangle$, $\mathsf{A}_2 = \langle\{\mathsf{a}, \mathsf{c}\}, \mathsf{a} \wedge \mathsf{c}\rangle$,
$\mathsf{C} = \langle\{\mathsf{c}\}, \mathsf{c}\rangle$, $\mathsf{A}_3 = \langle\{\mathsf{b}, \mathsf{c}\}, \mathsf{b} \wedge \mathsf{c}\rangle$,
$\mathsf{X} = \langle\{\mathsf{a}, \mathsf{c}\}, \neg\mathsf{b}\rangle$.

Comparing relevant arguments, we find: $\mathsf{A}_2 \lhd_{DE} \mathsf{A}_3$; $\mathsf{X} \lhd_{DE} A_3$; and also: $A_2 \lhd_{DD} A_3$; $\mathsf{X} \lhd_{DD} A_3$. Consequently, regarding both defeat relations $\hookrightarrow_{DE}$ and $\hookrightarrow_{DD}$ (with respect to the orderings $\lhd_{DE}$ and $\lhd_{DD}$, respectively), the same argument framework $(\mathit{Args}, \hookrightarrow)$ (where $\hookrightarrow = \hookrightarrow_{DE} = \hookrightarrow_{DD}$) results. It can be represented graphically as follows (in this and further illustrations of argument frameworks, nodes hold arguments and directed edges indicate defeats).



The sets $E = \{\mathsf{B}, \mathsf{C}, \mathsf{A}_3\}$ and $E' = \{\mathsf{A}, \mathsf{B}, \mathsf{A}_1\}$ are the only stable as well as preferred extensions of $(\mathit{Args}, \hookrightarrow)$. They represent that one can choose either Brownie and Cheesecake (E), or Almond cake and Brownie (E'). This is the 'rational' solution.

2.2 ASPIC$^+$

We turn to a well established structured argumentation formalism ASPIC$^+$ [16, 40, 41]. It is a widely used expressive approach representative of the use of *defeasible* rules[6] and preference aggregation in argumentation (see e.g. [26, 39, 40, 41, 47]).

ASPIC$^+$ can accommodate a wide range of instantiating logics, as well as preference relations. For our purposes, a simplified exposition of technical details will suffice: Cakes problem does not require use of defeasible rules, only strict rules; likewise, only one component among the premises (representing defeasible, or tentative, information) and axioms (representing facts) is needed, namely the premises. The details follow below.

Assume a logical language $\mathcal{L}$ closed under negation $\neg$. For $\varphi \in \mathcal{L}$, its *complement* $-\varphi$ is: $\neg\psi$ if $\varphi = \psi$; and ψ if $\varphi = \neg\psi$. To construct arguments, we will need to specify a set $\mathcal{R}_s$ of (strict) *rules* of the form $\varphi_1, \ldots, \varphi_n \to \psi$, where $\varphi_1, \ldots, \varphi_n, \psi \in \mathcal{L}$, and a set $\mathcal{K}_p \subseteq \mathcal{L}$ of *premises*. ASPIC$^+$ frameworks are meant to be normatively rational in the sense of [16], and in order to satisfy those rationality criteria, *contraposition* on rules is imposed [16, 41]:[7]

whenever $\varphi_1, \ldots, \varphi_n \to \psi \in \mathcal{R}_s$, we have $\varphi_1, \ldots, \varphi_{i-1}, -\psi, \varphi_{i+1}, \ldots, \varphi_n \to -\varphi_i \in \mathcal{R}_s$ too, for all i.

Arguments are then constructed inductively as deductions from premises using rules:

- if $\mathsf{a} \in \mathcal{K}_p$, then $[\mathsf{a}]$ is an argument with:
  - conclusion $\mathit{conc}([\mathsf{a}]) = \mathsf{a}$,

[5] By definition, there are infinitely many arguments, but it suffices to consider only a finite number of them, as they represent the essential information; see [11] for details.

[6] Two types of rules are commonly used in argumentation: *strict* rules, whose consequent necessarily follows from the antecedent; and *defeasible* rules, whose consequent *normally* (e.g. unless there are exceptions to the rule) follows from the antecedent.

[7] The principle that follows is in fact called *transposition* [16], while contraposition has a different formulation. Nonetheless, in ASPIC$^+$ the two are used interchangeably, and we chose the more appropriate one.

  - premises $prem([\mathsf{a}]) = \{\mathsf{a}\}$,
  - sub-arguments $sub([\mathsf{a}]) = \{[\mathsf{a}]\}$;

- if $\mathsf{A}_1, \ldots, \mathsf{A}_k$ are arguments with respective conclusions $conc(\mathsf{A}_1), \ldots, conc(\mathsf{A}_k)$, and there is a rule $conc(\mathsf{A}_1), \ldots, conc(\mathsf{A}_k) \to \mathsf{b} \in \mathcal{R}_s$, then $\mathsf{B} = [\mathsf{A}_1, \ldots, \mathsf{A}_k \to \mathsf{b}]$ is an argument with:
  - conclusion $conc(\mathsf{B}) = \mathsf{b}$,
  - premises $prem(\mathsf{B}) = prem(\mathsf{A}_1) \cup \ldots \cup prem(\mathsf{A}_k)$,
  - sub-arguments $sub(\mathsf{B}) = sub(\mathsf{A}_1) \cup \ldots \cup sub(\mathsf{A}_k) \cup \{\mathsf{B}\}$.

Attacks are obtained by deducing complements of premises:

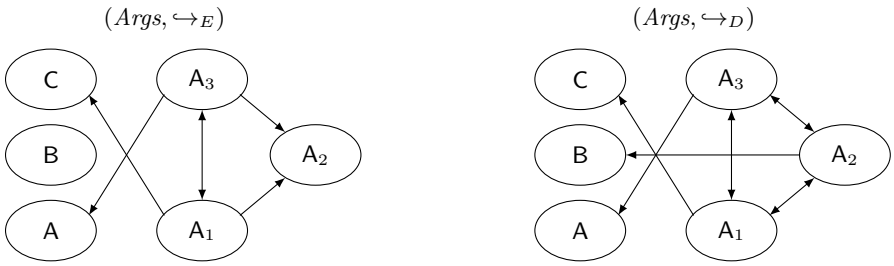
$\mathsf{A} \rightsquigarrow \mathsf{B}$ just in case $conc(\mathsf{A}) = -\mathsf{b}$ for some $\mathsf{b} \in prem(\mathsf{B})$.

An attack $\mathsf{A} \rightsquigarrow \mathsf{B}$ (on some $[\mathsf{b}] \in sub(\mathsf{B})$) succeeds as a defeat only if $\mathsf{A} \not\lhd [\mathsf{b}]$.[8] Even though in principle ASPIC$^+$ is completely modular with respect to the argument ordering principles to be used, certain requirements need to be satisfied $\lhd$ in order for ASPIC$^+$ frameworks to exhibit various desirable properties. In particular, ASPIC$^+$ offers two particular argument ordering principles, namely Elitist and Democratic, denoted by $\lhd_E$ and $\lhd_D$ respectively. For our purposes, given a preference relation $<$ over premises, they can be defined as follows: for $S, S' \subseteq \mathcal{K}_p$,

$S \lhd_E S'$ iff $\exists \mathsf{s} \in S$ such that $\forall \mathsf{s}' \in S'$ it holds that $\mathsf{s} < \mathsf{s}'$;
$S \lhd_D S'$ iff $\forall \mathsf{s} \in S$ it holds that $\exists \mathsf{s}' \in S'$ such that $\mathsf{s} < \mathsf{s}'$.

We next formalize the Cakes example in ASPIC$^+$.

Example 3 (Cakes in ASPIC$^+$)**.** Assuming language $\mathcal{L} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \neg\mathsf{a}, \neg\mathsf{b}, \neg\mathsf{c}\}$, we have the (strict) rules $\mathcal{R}_s = \{\mathsf{a}, \mathsf{c} \to \neg\mathsf{b}, \quad \mathsf{a}, \mathsf{b} \to \neg\mathsf{c}, \quad \mathsf{b}, \mathsf{c} \to \neg\mathsf{a}\}$ (note that closure under contraposition is imposed, so that there are three rules, rather than one) and premises $\mathcal{K}_p = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ (which stand for the assumed choices of cakes), together with preference $\mathsf{a} < \mathsf{b}$ of Brownie over Almond cake. We can construct arguments $\mathsf{A} = [\mathsf{a}]$, $\mathsf{B} = [\mathsf{b}]$, $\mathsf{C} = [\mathsf{c}]$, $\mathsf{A}_1 = [\mathsf{A}, \mathsf{B} \to \neg\mathsf{c}]$, $\mathsf{A}_2 = [\mathsf{A}, \mathsf{C} \to \neg\mathsf{b}]$, and $\mathsf{A}_3 = [\mathsf{B}, \mathsf{C} \to \neg\mathsf{a}]$, and obtain the following attacks: $\mathsf{A}_1 \rightsquigarrow \mathsf{C}, \mathsf{A}_2, \mathsf{A}_3$; $\mathsf{A}_2 \rightsquigarrow \mathsf{B}, \mathsf{A}_1, \mathsf{A}_3$; $\mathsf{A}_3 \rightsquigarrow \mathsf{A}, \mathsf{A}_1, \mathsf{A}_2$. Comparing argument premises yields (among other relationships) $\mathsf{A}_2 \lhd_E \mathsf{B}$, but $\mathsf{A}_2 \not\lhd_D \mathsf{B}$. Hence, with respect to the Elitist comparison, we find the following: $\mathsf{A}_2 \not\hookrightarrow_E \mathsf{B}$, A_1, A_3, whereas under the Democratic comparison, we find $\mathsf{A}_2 \hookrightarrow_D \mathsf{B}$, A_1, A_3. The other attacks succeed as defeats (under both comparison principles). Therefore, under the Elitist comparison, $\{\mathsf{B}, \mathsf{C}, \mathsf{A}_3\}$ and $\{\mathsf{A}, \mathsf{B}, \mathsf{A}_1\}$ are stable/preferred extensions. They represent the 'rational' choice. Under the Democratic comparison, $\{\mathsf{B}, \mathsf{C}, \mathsf{A}_3\}$, $\{\mathsf{A}, \mathsf{B}, \mathsf{A}_1\}$ and $\{\mathsf{A}, \mathsf{C}, \mathsf{A}_2\}$ are stable/preferred extensions. The 'rational' choice is reflected in the first two, while the third one represents an additional, and arguably unintuitive choice. The two argument frameworks are drawn below.



2.3 p_ABA

We next investigate a formalism called p_ABA [45], which is an extension of the well studied structured argumentation formalism Assumption-Based Argumentation (ABA) [12, 44]. While ABA itself does not possess a mechanism to deal with explicit preferences, p_ABA does. What is more, to the best of our knowledge, p_ABA is the only structured argumentation formalism to use preferences on the *extension level* to discriminate among extensions, and thus deserves a discussion. We start by giving details on ABA, based on [44], and then complement them with the ones necessary for p_ABA.

An ABA framework is a tuple $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$, where:

- $(\mathcal{L}, \mathcal{R})$ is a deductive system with a language (a set of sentences) $\mathcal{L}$ and a set $\mathcal{R}$ of rules of the form $\varphi_0 \leftarrow \varphi_1, \ldots, \varphi_m$ with $m \geqslant 0$ and $\varphi_i \in \mathcal{L}$ for $i \in \{0, \ldots, m\}$;
  - φ_0 is referred to as the *head* of the rule, and

[8] This definition of defeat in ASPIC$^+$ fixes a small non-essential error in the workshop version [20] of this paper.

  - $\varphi_1, \ldots, \varphi_m$ is referred to as the *body* of the rule;
  - if $m = 0$, then the rule $\varphi_0 \leftarrow \varphi_1, \ldots, \varphi_m$ is written as $\varphi_0 \leftarrow \top$, where $\top \notin \mathcal{L}$, and is said to have an empty body;
- $\mathcal{A} \subseteq \mathcal{L}$ is a non-empty set, whose elements are referred to as *assumptions*;
- $^- : \mathcal{A} \to \mathcal{L}$ is a total map: for $\alpha \in \mathcal{A}$, the $\mathcal{L}$-sentence $\overline{\alpha}$ is referred to as the *contrary* of α.

For the sake of simplicity, it suffices to focus on the so-called *flat* ABA frameworks, where no assumption is the head of any rule.

Arguments can be defined in terms of deductions. A *deduction for $\varphi \in \mathcal{L}$ supported by $S \subseteq \mathcal{L}$ and $R \subseteq \mathcal{R}$*, denoted by $S \vdash^R \varphi$, is a finite tree with

- the root labelled by φ,
- leaves labelled by $\top$ or elements from S,
- the children of non-leaf nodes ψ labelled by the elements of the body of some rule from $\mathcal{R}$ with head ψ, and R being the set of all such rules.

We say that there is an argument $\mathsf{A} : \ A \vdash \varphi$, whenever there is a deduction $A \vdash^R \varphi$, for some $R \subseteq \mathcal{R}$. For an argument $\mathsf{A} : \ A \vdash \varphi$, its *conclusion* $Cn(\mathsf{A})$ is φ. This naturally extends to sets E of arguments so that $Cn(E) = \{Cn(\mathsf{A}) \ : \ \mathsf{A} \in E\}$. Attacks are then defined via deductions of contraries of assumptions:

$\mathsf{A} : \ A \vdash \varphi \ \rightsquigarrow \ \mathsf{B} : \ B \vdash \psi$ just in case $\varphi = \overline{\mathsf{b}}$ for some $\mathsf{b} \in B$.

p_ABA frameworks $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \preccurlyeq)$ generalize ABA frameworks by adding a preference ordering $\preccurlyeq$ over $\mathcal{L}$. This ordering then induces an ordering $\sqsubseteq$ over extensions of the underlying ABA framework. For two extensions E and E' of $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$, $E \sqsubseteq E'$ essentially requires that there is a consequence φ of E' but not E, for which among consequences of E but not E' there is ψ but not χ with $\psi \preccurlyeq \varphi \prec \chi$ (where the strict counterpart $\prec$ of $\preccurlyeq$ is defined by $\alpha \prec \beta$ iff $\alpha \preccurlyeq \beta$ and $\beta \not\preccurlyeq \alpha$). Formally:

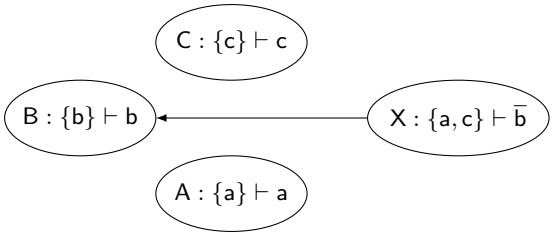
- $E \sqsubseteq E'$ if there is $\varphi \in Cn(E') \setminus Cn(E)$ such that
  - there is $\psi \in Cn(E) \setminus Cn(E')$ with $\psi \preccurlyeq \varphi$ and
  - there is no $\chi \in Cn(E) \setminus Cn(E')$ with $\varphi \prec \chi$;
- $E \sqsubseteq E$;
- if $E \sqsubseteq E'$ and $E' \sqsubseteq E''$, then $E \sqsubseteq E''$.

The latter two requirements ensure that $\sqsubseteq$ is a partial preorder, just like $\preccurlyeq$.

p_ABA then selects those extensions E that are not strictly less preferred than any other extension E' (i.e. $E \not\sqsubset E'$ for any E'), subsequently called $\mathcal{P}$-extensions.

Our Cakes example can be cast in p_ABA as follows.

Example 4 (Cakes in p_ABA)**.** We have a p_ABA framework $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \preccurlyeq)$ with language $\mathcal{L} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \overline{\mathsf{a}}, \overline{\mathsf{b}}, \overline{\mathsf{c}}\}$, rules $\mathcal{R} = \{\overline{\mathsf{b}} \leftarrow \mathsf{a}, \mathsf{c}\}$, assumptions $\mathcal{A} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ (standing for cakes) and preference $\mathsf{a} \prec \mathsf{b}$. We construct arguments $\mathsf{A} : \{\mathsf{a}\} \vdash \mathsf{a}$, $\mathsf{B} : \{\mathsf{b}\} \vdash \mathsf{b}$, $\mathsf{C} : \{\mathsf{c}\} \vdash \mathsf{c}$, and $\mathsf{X} : \{\mathsf{a}, \mathsf{c}\} \vdash \overline{\mathsf{b}}$. The only attack is $\mathsf{X} \rightsquigarrow B$, so the underlying ABA framework $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$, as depicted below, has a unique stable/preferred extension $\{\mathsf{A}, \mathsf{C}, \mathsf{X}\}$.



Since the ABA extension is unique, in p_ABA preferences do not play a role and $\{\mathsf{A}, \mathsf{C}, \mathsf{X}\}$ is a unique stable/preferred $\mathcal{P}$-extension of $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \preccurlyeq)$. This represents the arguably undesirable choice of Almond Cake and Cheesecake.

2.4 Abstract Argumentation

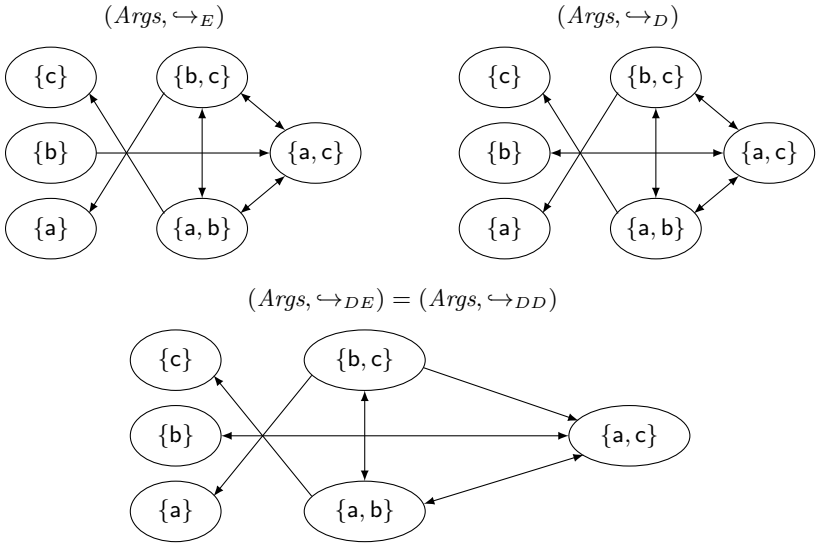
We finally consider AA with preferences. While originally ([25]) it did not incorporate preference handling mechanism, there are many proposals on how to extend AA with preference information, e.g. [1, 3, 5, 7, 14, 28, 34, 38]. Those approaches follow essentially the same guideline that an attack succeeds only if the attacker is not less preferred than the attackee. We therefore employ, following e.g. [1], a generic representative approach where given an AA framework $(Args, \rightsquigarrow)$ and a preference relation $\lhd$ over arguments, the defeat relation $\hookrightarrow$ is obtained from the condition $\mathsf{A} \hookrightarrow \mathsf{B}$ iff $\mathsf{A} \rightsquigarrow \mathsf{B}$ and $\mathsf{A} \not\lhd \mathsf{B}$.

Consider our attempt to render the Cakes problem in AA with preferences.

Example 5 (Cakes in AA)**.** Let a, b and c stand for taking a piece of Almond cake, the Brownie and a piece of Cheesecake, respectively. Every combination of a, b and c thus represents a choice of cakes, e.g. $\{\mathsf{b}, \mathsf{c}\}$ typifies the choice of Brownie and Cheesecake. With the attack relation specified next, we can simplify the argument framework so that it is easier to read: we can disregard the choice $\{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$, since taking all three pieces of cakes at once is not permitted; similarly, not taking anything, i.e. $\emptyset$, can be discarded. So let $Args = \wp(\{\mathsf{a}, \mathsf{b}, \mathsf{c}\}) \setminus \{\emptyset, \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}\} = \{\{\mathsf{a}\}, \{\mathsf{b}\}, \{\mathsf{c}\}, \{\mathsf{a}, \mathsf{b}\}, \{\mathsf{a}, \mathsf{c}\}, \{\mathsf{b}, \mathsf{c}\}\}$ be the set of arguments.

We can then stipulate the attack relation as follows: for $\mathsf{S}, \mathsf{S}' \in Args$, let $\mathsf{S} \rightsquigarrow \mathsf{S}'$ stand for 'if you want the cakes from S, do not choose S''. For example, if Brownie and Cheesecake is wanted, then do not take Almond cake and Cheesecake, and vice versa, i.e. $\{\mathsf{b}, \mathsf{c}\} \rightsquigarrow \{\mathsf{a}, \mathsf{c}\}$ and $\{\mathsf{a}, \mathsf{c}\} \rightsquigarrow \{\mathsf{b}, \mathsf{c}\}$. Similarly, if you want Brownie and Cheesecake, then do not choose Almond cake, i.e. $\{\mathsf{b}, \mathsf{c}\} \rightsquigarrow \{\mathsf{a}\}$; yet if Almond cake is desired, then it is not a problem to take Brownie and Cheesecake, i.e. $\{\mathsf{a}\} \not\rightsquigarrow \{\mathsf{b}, \mathsf{c}\}$. This way, due to the rule that taking Almond cake and Cheesecake prevents getting Brownie, we have (among others) an attack $\{\mathsf{b}\} \rightsquigarrow \{\mathsf{a}, \mathsf{c}\}$.

Preference information tells us that $\mathsf{a} < \mathsf{b}$. To account for this we can employ the Elitist and Democratic ordering principles, as well as their Disjoint counterparts, defined in Subsect.s 2.2 and 2.1. Then $\mathsf{S} \hookrightarrow_X \mathsf{S}'$ iff $\mathsf{S} \rightsquigarrow \mathsf{S}'$ and $\mathsf{S} \not\lhd_X S'$, where $X \in \{E, D, DE, DD\}$. Relevant to the attacks, we find: $\{\mathsf{a}, \mathsf{c}\} \lhd_E \{\mathsf{b}\}$; $\{\mathsf{a}, \mathsf{c}\} \lhd_{DE} \{\mathsf{b}, \mathsf{c}\}$; $\{\mathsf{a}, \mathsf{c}\} \lhd_{DD} \{\mathsf{b}, \mathsf{c}\}$. So we obtain that $\{\mathsf{a}, \mathsf{c}\} \not\hookrightarrow_E \{\mathsf{b}\}$, $\{\mathsf{a}, \mathsf{c}\} \not\hookrightarrow_{DE} \{\mathsf{b}, \mathsf{c}\}$ and $\{\mathsf{a}, \mathsf{c}\} \not\hookrightarrow_{DD} \{\mathsf{b}, \mathsf{c}\}$. The following graphically depicts the argument frameworks in question.



In $(Args, \hookrightarrow_E)$, $(Args, \hookrightarrow_{DE})$ and $(Args, \hookrightarrow_{DD})$, the sets $E = \{\{\mathsf{b}\}, \{\mathsf{c}\}, \{\mathsf{b}, \mathsf{c}\}\}$ and $E' = \{\{\mathsf{a}\}, \{\mathsf{b}\}, \{\mathsf{a}, \mathsf{b}\}\}$ are the only stable/preferred extensions. They correspond to the 'rational' choice. Meanwhile, $(Args, \hookrightarrow_D)$, in addition to E and E', admits $E'' = \{\{\mathsf{a}\}, \{\mathsf{c}\}, \{\mathsf{a}, \mathsf{c}\}\}$ as a stable/preferred extension, which corresponds to the undesirable Almond Cake and Cheesecake combination.

Recently, another type of modification of attacks in AA due to preferences was proposed: in *Preference-based Argumentation Frameworks* (PAFs) [3], the preference relation $\lhd$ over arguments is used to *reverse*, rather than discard attacks. Such an approach supposedly avoids several issues arising when attacks are merely removed due to preference information, see [3]. However, in Example 5 such attack reversal would result into already existing attacks (because the relevant pairs of arguments already symmetrically attack each other), so the outcome would be the same. Nonetheless, the idea of attack reversal can be exploited in structured argumentation, with different results. Indeed, we next briefly discuss a very recently proposed formalism, *Assumption-Based Argumentation with Preferences* (ABA$^+$) [21], that illustrates the use of attack reversal in structured argumentation and is the only approach investigated in this paper that does opt for the 'intuitive' solution in Cakes problem.

2.5 ABA$^+$

ABA$^+$ [21, 22] is a formalism generalizing ABA (briefly presented in Subsect. 2.3), that equips ABA frameworks with a preference ordering (transitive binary relation) $\leqslant$ on assumptions $\mathcal{A}$. The preferences are incorporated directly into the attack relation, thus avoiding preference aggregation (cf. Subsect.s 2.1 and 2.2) and comparison among extensions (cf. Subsect. 2.3). The idea of attack reversal is that when the attacker has an assumption less preferred than the one attacked, then the attack is reversed. Also, instead of constructing arguments as such, in ABA$^+$ we work directly with sets of assumptions (this can be done in ABA and p_ABA too, see e.g. [27]), so the notion of argument can be dropped altogether. Formal details are as follows.

An ABA$^+$ framework is a tuple $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leqslant)$, where $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ is an ABA framework and $\leqslant$ is a transitive binary relation on $\mathcal{A}$. A set $A \subseteq \mathcal{A}$ of assumptions *<-attacks* a set $B \subseteq \mathcal{A}$ of assumptions, denoted $A \rightsquigarrow_< B$, just in case (see Subsect. 2.3 for the definition of deductions in ABA):

- either there is a deduction $A' \vdash^R \overline{\beta}$, for some $\beta \in B$, supported by $A' \subseteq A$, and $\nexists \alpha' \in A'$ with $\alpha' < \beta$;
- or there is a deduction $B' \vdash^R \overline{\alpha}$, for some $\alpha \in A$, supported by $B' \subseteq B$, and $\exists \beta' \in B'$ with $\beta' < \alpha$.

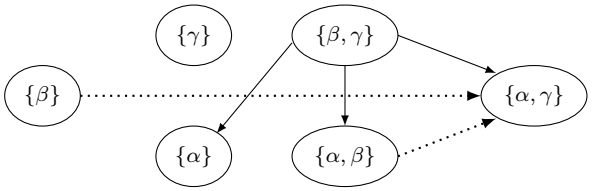
The first type of attack is called *normal*, and the second one *reverse*.

In other words, if $A \rightsquigarrow B$ and no assumption of A used in this attack is strictly less preferred than the attacked assumption (from B), then we obtain $A \rightsquigarrow_< B$ as a normal attack. Otherwise, if $A \rightsquigarrow B$ and this attack depends on at least one assumption that is strictly less preferred than the attacked one, then the attack is reversed and we obtain $B \rightsquigarrow_< A$ instead.

ABA$^+$ semantics are then defined as in Definition 1, but with arguments and membership replaced by sets of assumptions and subset inclusion, respectively. (A prefix <- is added, e.g. <-preferred, to differentiate between ABA and ABA$^+$ semantics.) Finally, similarly to ASPIC$^+$, a condition on the rules of a framework can be imposed, namely a relaxed version of contraposition, which states that contraposing is needed only when a deduction involves assumptions less preferred than the one whose contrary is deduced.[9]

Let us see how ABA$^+$ copes with our Cakes example.

Example 6 (Cakes in ABA$^+$)**.** Let $\mathcal{R} = \{\overline{\beta} \leftarrow \alpha, \gamma, \quad \overline{\alpha} \leftarrow \beta, \gamma\}$, $\mathcal{A} = \{\alpha, \beta, \gamma\}$ (standing for Almond Cake, Brownie and Cheesecake, respectively) and $\alpha < \beta$. (The language and contrary are implicit from $\mathcal{A}$ and $\mathcal{R}$.) The presence of the rule $\overline{\alpha} \leftarrow \beta, \gamma$ ensures that the relaxed version of contraposition mentioned above is satisfied, by guaranteeing that the deduction $\{\alpha, \gamma\} \vdash^{\{\overline{\beta} \leftarrow \alpha, \gamma\}} \overline{\beta}$, supported by the assumption α which is less preferred than the assumption β whose contrary is deduced, 'contraposes'. Observe as well that (full) contraposition is still violated (e.g. missing $\overline{\gamma} \leftarrow \alpha, \beta$)[10]. The framework is illustrated below (the irrelevant empty set $\emptyset$, the <-self-attacking $\mathcal{A}$ as well as <-attacks to and from $\mathcal{A}$ are omitted for readability):



Here, the set $\{\beta, \gamma\}$ of assumptions is a unique <-preferred as well as <-stable extension. It represents the 'intuitive' choice of Brownie and Cheesecake.

3 Related Work and Discussion

We briefly discuss a few distinguished formalisms, namely Extended Argumentation Frameworks (EAFs) [38, 39], Argumentation Frameworks with Recursive Attacks (AFRA) [5] and Defeasible Logic Programming (DeLP) [29, 30]. Both EAFs and AFRA are generalizations of AA that allow for attacks to be directed at attacks, as well as at arguments. The intuition is that an attack on an attack $\mathsf{A} \rightsquigarrow \mathsf{B}$ expresses a preference for B over A. However, if a preference relation among arguments is not explicitly given, as in the Cakes example, then it is not trivial to figure out what additional attacks on attacks should be introduced. In any event, the effect of an attack directed at another attack is to disable the latter one, and thus one can say that the approaches fall somewhat into the category discussed in Subsect. 2.4.

[9] This relaxed version of contraposition has not been published or presented as part of the ABA$^+$ formalism yet, wherefore we do not give a formal statement and use it informally in the example that follows.

[10] Technically, regarding contraries as complements.

DeLP, on the other hand, is another structured argumentation formalism that relies on defeasible rules (see Subsect. 2.2), and in addition involves the concept of negation as failure [19] from Logic Programming.[11] To represent any non-factual information DeLP employs defeasible rules, and if negation as failure is used, then attacks succeed irrespective of the preferences, like in ABA (cf. Subsect. 2.3).

In terms of modelling the Cakes problem from Example 1 in the formalisms considered, we concede that different representations are possible. We chose a natural way to represent a single (strict) rule and a single preference involving two alternatives, and believe that other representations would come at the expense of perspicuity and generality.

We have seen that among the selected formalisms of argumentation (both abstract and structured) with preferences almost none opt exclusively for the 'intuitive' answer {Brownie, Cheesecake} (as supported by the survey results) when modelling the Cakes problem. In Deductive Argumentation (Sect. 2.1) the normative 'rational' solution of choosing indifferently between {Almond cake, Brownie} and {Brownie, Cheesecake} was obtained. ASPIC$^+$ (Sect. 2.2) yielded other choices (such as {Almond cake, Cheesecake}) in addition to the 'rational' one. Preference information in p_ABA (Sect. 2.3) did not really play a role, and just like ABA, the formalism generated a unique choice {Almond cake, Cheesecake}. AA with preferences (Sect. 2.4) produced either the 'rational choice', or additionally the one afforded by p_ABA too. However, we finally encountered ABA$^+$, that did produce the 'intuitive' {Brownie, Cheesecake} solution.

The fact that most current approaches do not opt exclusively for the 'intuitive' solution in the Cakes example does not mean those argumentation formalisms are ill-suited to handle preferences. Each formalism has been shown to satisfy some desirable properties and adhere to certain intuitions. Some approaches may also aim for being more normative systems (e.g. ASPIC$^+$), in the sense of complying with certain rationality standards (such as [16]), rather than mimicking human intuition at the expense of rationality. Indeed, common-sense reasoning and human decision making are not necessarily rational endeavours (see e.g. [35]). However, as argumentation is a technique to model common-sense reasoning, it is natural to demand of formalisms that aspire to this goal to be able to accordingly model common-sense reasoning scenarios such as the one considered in this paper. We are therefore hopeful that ABA$^+$ will at least be able to come closer to fulfilling this goal.

4 Conclusion and Future Work

In this paper we investigated whether and how existing approaches to argumentation with preferences model human intuition behind common-sense reasoning scenarios involving rules and preference information. To this end, we proposed an example of a decision making problem that assumes essentially a single strict rule and a single preference. The example was supplemented with both an informal solution to the problem, referred to as the 'rational' choice, and a survey, results of which indicated a solution, referred to as the 'intuitive' choice, that human respondents have opted for most often. We rendered the problem in several representative formalisms of argumentation with preferences, and saw that all of them but one produced either the 'rational' or some other choices, but not exclusively the 'intuitive' one. The only one to opt exclusively for the 'intuitive' solution was the recently proposed structured argumentation formalism that utilizes attack reversal, namely ABA$^+$ [21]. As argumentation is used in areas such as common-sense reasoning and decision making among others, the results, we argue, call for a deeper analysis of how to adequately represent in argumentation the type of problems as the one considered here. Still further, our investigation opens up space for discussions on how to deal with such reasoning problems involving preferences in argumentation-based approaches to, for instance, multi-agent systems, as well as in AI more generally. We plan to pursue research regarding these general problems further.

Additionally, we plan to study scenarios as in the Cakes example themselves and their treatment in AI. In particular, it would be interesting to see what existing reasoning techniques (e.g. constraint logic programming, Event Calculus) would capture human intuitions in such scenarios. This research direction also includes enquiry into the meaning of preferences and their relation to, for example, rules. In addition, it involves exploring what formal properties regarding preferences can be conceived, as in e.g. [13]. It furthermore means designing and running more surveys to get a better grasp behind human intuitions. Such research is essential to understanding preferences and approaching common-sense reasoning problems from the argumentation perspective in particular, and from the point of view of AI in general.

There are known issues concerning technical points, theoretical appropriateness as well as intuitiveness of different approaches to argumentation with preferences, see e.g. [3, 6, 17, 32, 40]. In AI at large there is much debate on how to manage preferences (see e.g. [24, 33]). Our work contributes to these lines of research by presenting a simple example of reasoning with preferences, opening up space for discussions on how to cope with such common-sense reasoning scenarios in argumentation, and challenging the existing and potential argumentation formalisms to model the human intuition behind them.

[11] In fact, ABA (Subsect. 2.3) is also heavily influenced by Logic Programming.

References

[1] Leila Amgoud and Claudette Cayrol. A Reasoning Model Based on the Production of Acceptable Arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002. doi:10.1023/A:1014490210693.

[2] Leila Amgoud and Henri Prade. Using Arguments for Making and Explaining Decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009. doi:10.1016/j.artint.2008.11.006.

[3] Leila Amgoud and Srdjan Vesic. Rich Preference-Based Argumentation Frameworks. *International Journal of Approximate Reasoning*, 55(2):585–606, 2014. doi:10.1016/j.ijar.2013.10.010.

[4] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An Introduction to Argumentation Semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011. doi:10.1017/S0269888911000166.

[5] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. AFRA: Argumentation Framework with Recursive Attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011. doi:10.1016/j.ijar.2010.05.004.

[6] Pietro Baroni, Massimiliano Giacomin, and Beishui Liao. I don't care, I don't know ... I know too much! On Incompleteness and Undecidedness in Abstract Argumentation. In Thomas Eiter, Hannes Strass, Mirosław Truszczyński, and Stefan Woltran, editors, *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation: Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, volume 9060 of *Lecture Notes in Computer Science*, pages 265–280. Springer, 2015. doi:10.1007/978-3-319-14726-0_18.

[7] Trevor J M Bench-Capon. Persuasion in Practical Argument Using Value Based Argumentation Frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003. doi:10.1093/logcom/13.3.429.

[8] Trevor J M Bench-Capon and Paul E Dunne. Argumentation in Artificial Intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007. doi:10.1016/j.artint.2007.05.001.

[9] Philippe Besnard, Alejandro Javier García, Anthony Hunter, Sanjay Modgil, Henry Prakken, Guillermo Ricardo Simari, and Francesca Toni. Introduction to Structured Argumentation. *Argument & Computation*, 5(1):1–4, 2014. doi:10.1080/19462166.2013.869764.

[10] Philippe Besnard and Anthony Hunter. A Logic-based Theory of Deductive Arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001. doi:doi:10.1016/S0004-3702(01)00071-6.

[11] Philippe Besnard and Anthony Hunter. Constructing Argument Graphs with Deductive Arguments: A Tutorial. *Argument & Computation*, 5(1):5–30, 2014. doi:10.1080/19462166.2013.869765.

[12] Andrei Bondarenko, Phan Minh Dung, Robert Kowalski, and Francesca Toni. An Abstract, Argumentation-Theoretic Approach to Default Reasoning. *Artificial Intelligence*, 93(97):63–101, 1997. doi:10.1016/S0004-3702(97)00015-5.

[13] Gerhard Brewka and Thomas Eiter. Preferred Answer Sets for Extended Logic Programs. *Artificial Intelligence*, 109(1-2):297–356, 1999. doi:10.1016/S0004-3702(99)00015-6.

[14] Gerhard Brewka, Stefan Ellmauthaler, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Abstract Dialectical Frameworks Revisited. In Francesca Rossi, editor, *23rd International Joint Conference on Artificial Intelligence*, pages 803–809, Beijing, 2013. IJCAI/AAAI.

[15] Gerhard Brewka, Ilkka Niemelä, and Mirosław Truszczyński. Preferences and Nonmonotonic Reasoning. *AI Magazine*, 29(4):69–78, 2008.

[16] Martin Caminada and Leila Amgoud. On the Evaluation of Argumentation Formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007. doi:10.1016/j.artint.2007.02.003.

[17] Martin Caminada and Yining Wu. On the Limitations of Abstract Argumentation. In *23rd Benelux Conference on Artificial Intelligence*, volume 23, pages 59–66, Ghent, 2011.

[18] Álvaro Carrera and Carlos Iglesias. A Systematic Review of Argumentation Techniques for Multi-Agent Systems Research. *Artificial Intelligence Review*, 44(4):509–535, 2015. doi:10.1007/s10462-015-9435-9.

[19] Keith Clark. Negation as Failure. In Hervé Gallaire and Jack Minker, editors, *Logic and Data Bases*, pages 293–322. Springer, 1978. doi:10.1007/978-1-4684-3384-5_11.

[20] Kristijonas Čyras. Argumentation-Based Reasoning with Preferences. In Javier Bajo, Maria José Escalona, Sylvain Giroux, Patrycja Hoffa-Dabrowska, Vicente Julián, Paulo Novais, Nayat Sánchez Pi, Rainer Unland, and Ricardo Azambuja Silveira, editors, *Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection - International Workshops of PAAMS 2016*, volume 616 of *Communications in Computer and Information Science*, pages 199–210, Sevilla, 2016. Springer. doi:10.1007/978-3-319-39387-2_17.

[21] Kristijonas Čyras and Francesca Toni. ABA+: Assumption-Based Argumentation with Preferences. In Chitta Baral, James P Delgrande, and Frank Wolter, editors, *Principles of Knowledge Representation and Reasoning, 15h International Conference*, pages 553–556, Cape Town, 2016. AAAI Press.

[22] Kristijonas Čyras and Francesca Toni. Properties of ABA+ for Non-Monotonic Reasoning. In *16th International Workshop on Non-Monotonic Reasoning*, pages 25–34, Cape Town, 2016.

[23] James P Delgrande, Torsten H Schaub, Hans Tompits, and Kewen Wang. A Classification and Survey of Preference Handling Approaches in Nonmonotonic Reasoning. *Computational Intelligence*, 20(2):308–334, 2004. doi:10.1111/j.0824-7935.2004.00240.x.

[24] Carmel Domshlak, Eyke Hüllermeier, Souhila Kaci, and Henri Prade. Preferences in AI: An overview. *Artificial Intelligence*, 175(7-8):1037–1052, 2011. doi:10.1016/j.artint.2011.03.004.

[25] Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games. *Artificial Intelligence*, 77:321–357, 1995. doi:10.1016/0004-3702(94)00041-X.

[26] Phan Minh Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016. doi:10.1016/j.artint.2015.10.005.

[27] Phan Minh Dung, Robert Kowalski, and Francesca Toni. Assumption-Based Argumentation. In Guillermo Ricardo Simari and Iyad Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009. doi:10.1007/978-0-387-98197-0_10.

[28] Paul E Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted Argument Systems: Basic Definitions, Algorithms, and Complexity Results. *Artificial Intelligence*, 175(2):457–486, 2011. doi:10.1016/j.artint.2010.09.005.

[29] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(2):95–138, 2004. arXiv:0302029, doi:10.1017/S1471068403001674.

[30] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible Logic Programming: DeLP-servers, Contextual Queries, and Explanations for Answers. *Argument & Computation*, 5(1):63–88, 2014. doi:10.1080/19462166.2013.869767.

[31] Nikos Gorogiannis and Anthony Hunter. Instantiating Abstract Argumentation with Classical Logic Arguments: Postulates and Properties. *Artificial Intelligence*, 175(9-10):1479–1497, 2011. doi:10.1016/j.artint.2010.12.003.

[32] Souhila Kaci. Refined Preference-Based Argumentation Frameworks. In Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo Ricardo Simari, editors, *Computational Models of Argument*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 299–310, Desenzano del Garda, 2010. IOS Press. doi:10.3233/978-1-60750-618-8-299.

[33] Souhila Kaci. *Working with Preferences. Less is More*. Springer, 2011. arXiv:arXiv:1011.1669v3, doi:10.1017/CBO9781107415324.004.

[34] Souhila Kaci and Leendert van der Torre. Preference-Based Argumentation: Arguments Supporting Multiple Values. *International Journal of Approximate Reasoning*, 48(3):730–751, 2008. doi:10.1016/j.ijar.2007.07.005.

[35] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291, 1979.

[36] Antonis C Kakas and Pavlos Moraitis. Argumentation Based Decision Making for Autonomous Agents. In *2nd International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 883–890, Melbourne, 2003. ACM Press. doi:10.1145/860575.860717.

[37] Hugo Mercier and Dan Sperber. Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, 34(2):57–74, 2011. doi:10.1017/S0140525X10000968.

[38] Sanjay Modgil. Reasoning About Preferences in Argumentation Frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009. doi:10.1016/j.artint.2009.02.001.

[39] Sanjay Modgil and Henry Prakken. Reasoning About Preferences in Structured Extended Argumentation Frameworks. In Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo Ricardo Simari, editors, *Computational Models of Argument*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 347–358, Desenzano del Garda, 2010. IOS Press. doi:10.3233/978-1-60750-619-5-347.

[40] Sanjay Modgil and Henry Prakken. A General Account of Argumentation with Preferences. *Artificial Intelligence*, 195:361–397, 2013. doi:10.1016/j.artint.2012.10.008.

[41] Sanjay Modgil and Henry Prakken. The ASPIC+ Framework for Structured Argumentation: A Tutorial. *Argument & Computation*, 5(1):31–62, 2014. doi:10.1080/19462166.2013.869766.

[42] Gabriella Pigozzi, Alexis Tsoukiàs, and Paolo Viappiani. Preferences in Artificial Intelligence. *Annals of Mathematics and Artificial Intelligence*, pages 1–41, 2015. doi:10.1007/s10472-015-9475-5.

[43] Iyad Rahwan and Guillermo Ricardo Simari. *Argumentation in Artificial Intelligence*. Springer, 2009. doi:10.1007/978-0-387-98197-0.

[44] Francesca Toni. A Tutorial on Assumption-Based Argumentation. *Argument & Computation*, 5(1):89–117, 2014. doi:10.1080/19462166.2013.869878.

[45] Toshiko Wakaki. Assumption-Based Argumentation Equipped with Preferences. In Hoa Khanh Dam, Jeremy V Pitt, Yang Xu, Guido Governatori, and Takayuki Ito, editors, *Principles and Practice of Multi-Agent Systems - 17th International Conference*, volume 8861 of *Lecture Notes in Computer Science*, pages 116–132, Gold Coast, 2014. Springer. doi:10.1007/978-3-319-13191-7_10.

[46] Toshiko Wakaki. Preference-Based Argumentation Built from Prioritized Logic Programming. *Journal of Logic and Computation*, 25(2):251–301, 2015. doi:10.1093/logcom/exs066.

[47] Anthony P Young, Sanjay Modgil, and Odinaldo Rodrigues. Prioritised Default Logic as Rational Argumentation. In Catholijn M Jonker, Stacy Marsella, John Thangarajah, and Karl Tuyls, editors, *16th International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 626–634, Singapore, 2016. ACM.